DOCUMENT RESUME

ED 305 387 TM 012 927

AUTHOR Rule, David L.; Grippin, Pauline C.
TITLE A Critical Comparison of Learning Style Instruments Frequently Used with Adult Learners.
PUB DATE Feb 88
NOTE 24p.; Paper presented at the Annual Conference of the Eastern Educational Research Association (Miami Beach, FL, February 24-27, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Adult Learning; Adults; *Cognitive Measurement; Cognitive Style; Comparative Analysis; Evaluation Methods; Learning Processes; *Learning Strategies; *Psychometrics; Test Construction; Test Interpretation; Test Reliability; *Test Reviews; Test Use; Test Validity
IDENTIFIERS Group Embedded Figures Test; Learning Style Inventory (Kolb); Myers Briggs Type Indicator; Personal Style Indicator; Productivity Environmental Preferences Survey; Self Directed Learning Readiness Scale

ABSTRACT

Theoretical framework, instrument development, and psychometric qualities are reviewed for some measures of learning style currently in use with adults. Instruments reviewed include: (1) Kolb's Learning Style Inventory I; (2) Kolb's Learning Style Inventory II; (3) the Myers-Briggs Type Indicator; (4) the Productivity Environmental Preferences Survey; (5) the Self-Directed Learning Readiness Scale; (6) the Personal Style Indicator; and (7) the Group Embedded Figures Test. For each review, the reliability, validity, interpretation, and utility of the test are discussed. Reviewed instruments come from various theoretical perspectives and have been in existence for varying lengths of time. The review does not indicate that older instruments are more reliable or valid than newer counterparts, but does demonstrate the vast amount of information available for older instruments. For instruments more recently developed, more quality information is needed for the user to make informed decisions about their use. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED305387

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Pauline C. Grippin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A Critical Comparison of Learning Style Instruments Frequently Used with Adult Learners

David L. Rule
State University of New York at Albany, Albany, New York

Pauline C. Grippin
The College of Saint Rose, Albany, New York

Paper presented at the annual conference of the Eastern Educational Research Association at Miami Beach, Florida, February 25, 1988

TM012927

Based on a review of literature, the construct of learning style appears to be broad and the instruments purported to assess it uneven. The choice of a learning style instrument has the potential for skewing experimental, correlational, descriptive, and guidance outcomes in ways that are not always obvious to the user. Instruments in current use by practitioners (researchers, teachers, counselors, and administrators) to provide guidance to adult learners and/or research phenomena of the adult learner pose problems. These problems include instrument development, reliability, and validity, therefore effecting the quality of both research results and guidance alike.

Perhaps adding to these problems is that the cognitive style/learning style continuum is not yet clearly defined. The construct of learning style is often used interchangeably with cognitive style, but often denotes a broader scope. Cognitive style refers to an individual's preferred mode of perceiving information and/or cognitive functioning, thus including both perceptual and intellectual functioning. Learning style, however, often encompasses an individual's preference for environmental factors found in the learning situation, as well as cognitive style. This lack of common definition has resulted in similar sounding measures actually coming from quite different theoretical perspectives. Also, the psychometric qualities of these learning style instruments vary from very weak to moderately strong and research studies present conflicting results. These two factors, lack of common definition and varying instrument quality, make each measure's usefulness for guidance of and/or research on adult learners problematic.

This paper will review the theoretical framework, instrument development, and psychometric qualities of six instruments in current use. Following each review are some conclusions about the interpretability/use-ability of individual scores. The instruments reviewed include: Kolb's Learning Style Inventory I and II (1976, 1985), Guglielmino and Guglielmino's Self-Directed Learning Readiness Scale (1977), Oltman, Raskin, & Witkin's Group Embedded Figures Test (1971), Price, Dunn, and Dunn's Productivity Environmental Preference Survey (1985), and Cotroneo's Personal Survey Indicator (1983). Also, the Myers-Briggs Type Indicator (1985).

KOLB'S LEARNING STYLE INVENTORY I & II (1976, 1985)

LEARNING STYLE INVENTORY I (1976)

Kolb's Learning Style Inventory I (LSI-I) (1976) is one of the most frequently cited learning style instruments in research examining adult learners.

The Technical Manual (Kolb, 1976) describes the original theoretical basis of the LSI I as experiential learning, an extension of Jungian and Lewinian concepts, delineating three stages of life: acquisition,

specialization, and integration. Cognitive development over these stages is mediated by a cyclical tension across two polarities utilizing four learning modes: concrete experience versus abstract conceptualization and active experimentation versus reflective observation. The LSI I purports to discriminate among learners on these two continua and to identify four types of learning style: Converger, Diverger, Assimilator, and Accomodator.

The Converger combines Abstract Conceptualization and Active Experimentation, enjoying the practical application of ideas. Kolb describes the Converger as unemotional and good at tasks requiring convergent results. The Diverger combines Concrete Experience and Reflective Abstraction, excelling in generating new wholes. This type is said to be imaginative, emotional and people-oriented. The Assimilator combines Abstract Conceptualization and Reflective Observation, preferring theory to people and inductive rather than deductive reasoning. The Accomodator combines Concrete Experience and Active Experimentation, wanting to do rather than think. This type relates to the immediate issue in a trial and error manner.

Instrument Development/Format

The initial stage of development utilized "four behavioral scientists acquainted with experiential learning theory" (Kolb, 1976, p.10) to generate lists of words that reflect the four described learning modes and final item selection was done by correlating each item to the total score for the learning mode subscale: concrete experience, abstract conceptualization, active experimentation, and reflective observation.

The sample used for the final item selection consisted entirely of managers and students in management programs.

The format is a 9-item self-report questionnaire (ipsative) in which the respondent rank orders four words per item with each word representing a learning mode. The respondent is asked to rank these four words in a way that best describes his/her learning style. A raw score is then obtained for each of the four learning modes: concrete experience, abstract conceptualization, active experimentation, and reflective observation. Learning style types are identified by calculating the difference scores of the two mode polarities and then plotting the difference scores on a graph having the two continua at right angles. The intersection of the difference scores place the learner in a learning style's quadrant: Converger, Diverger, Assimilator, and Accomodator.

Reliabilities

The internal reliability of the LSI I is problematic. There is interdependence of items since the constructs constitute a continuum and the dimensions have been set up

as two polarities. Reported internal consistency coefficients for the four dimensions, based on three items each, are mostly in the low .60's (Kolb, 1976). The correlations of the difference scores between Abstract Conceptualization and Concrete Experience are in the high .70's and between Active Experimentation and Reflective Observation are in the low .80's (Kolb, 1976). However, a more recent study by Sims, Veres III, Watson, and Buckner (1986) reports internal consistencies for the four scales (coefficient alpha) ranging to from .23 (Active Experimentation) to .58 (Reflective Observation). Merritt and Marshall (1984, as cited in Sewall, 1986) also report low internal consistency (alpha) ranging from .29 (Concrete Experience) to .59 (Reflective Observation).

Kolb (1976) reports stability coefficients, for the four learning modes, ranging from the .60's after three months to the .40's after seven months. The combined (difference) scores, between Abstract Conceptualization and Concrete Experience and between Active Experimentation and Reflective Observation, display a similar range. However, authors report stability coefficients for the four scales ranging from .52 to .61 after 31 days (Geller, 1979), .39 to .63 after 5 weeks (Freedman & Stumpf, 1978), and .45 to .53 after ten weeks (Sims et al., 1986).

Validity

The validity evidence offered by the Technical Manual (1976) is not strong. Scores for various groups do show variability. However, the standard deviations of the difference scores are considerably higher than the standard deviations of the separate raw scores. Given the concern for the LSI's reliability, the estimated standard error of measurement (not reported) places in jeopardy the interpretation/utilization of difference scores when determining the individual's learning style.

The management sample used in the final item selection process also threatens validity. If the theoretical framework of the LSI I is correct, individuals who are attracted to a common profession and tested at similar developmental stages would be expected to share certain characteristics. Using such a homogeneous sample for final item selection could introduce unknown bias into the instrument. This concern is supported by a study which performed a principal component analysis of the LSI and found that of the four obtained factors, only one resembled a dimension theorized by Kolb (Wunderlich & Gjerde, 1978). This study's sample was, like the sample used for LSI development, rather homogeneious, utilizing fourth year medical school student and practicing M.D.s.

In a more recent study by Holtzclaw (1985), the relationships between learning styles, choice of courses, and experiental learning credit were examined. Holtzclaw (1985) reports that the congruences between the Kolb model

and his results were moderate for Divergers, Convergers, and Accommondators and totally incongruent for Assimilators.

Interpretation

The LSI I's internal consistency appears to be in doubt. Although the theoretical basis for the instrument would support changes in learning style over developmental stages, such low internal consistency and stability reliabilities over such short time frames threatens the instrument's validity. Also, in addition to the concerns for validity stated above, other researchers looking at the viability of the LSI have found it to be unreliable and/or invalid (West, 1982; Fox, 1984; Freedman & Stumpf, 1978; Korhonen & McCall, 1986; Marshall & Merritt, 1985; Pigg, Busch, & Lacy, 1980; Sugarman, 1985).

Given the conflicting reports of reliability, between the Technical Manual (1976) and research results, and the conflicting support for validity of the LSI I, interpretation of individual scores should be done with reference to the normative data supplied by the LSI I and studies utilizing similar populations. If such normative data is unavailable, a pilot study using a sample from the population of interest is warranted. Further, given the concerns over the instrument's reliability, individual difference scores near the intersection of the learning mode continua should not be interpretated as being definitive.

LEARNING STYLE INVENTORY II (1985)

The LSI II (1985) is a revised version of the LSI-I, having the same theoretical basis, but including "improvements designed to enhance the scientific measurement specifications and the inventory's practical uses in education and counselling" (1985, p.1).

Instrument Development/Format

The format of the LSI has changed with the revised version; LSI II (Technical Specifications, 1985). The revised LSI II has three more items than the original LSI, bringing the total number of items to 12. Some of the LSI II's items were drawn from the original LSI I (1975), "revised into simpler language" (1985, p.4) and six new items were created per scale. Difference scores between learning modes are maintained in the calculation to determine an individual's learning style. The ipsative item format is also maintained and revised reliability coefficients and normative samples are presented.

Reliabilities

The LSI II (1985) reports internal consistency coefficients (Cronbach's alpha) ranging from .73 (Reflective

Observation) to .83 (Abstract Conceptualization) for the four learning modes. Difference score internal consistencies are reported as .88 (Abstract-Concrete) and .81 (Active-Reflective). These improvements over the LSI I (1975) internal consistencies are supported by a study (Sims et al., 1986) which reports alpha's ranging from .76 (Concrete Experience) to .85 (Abstract Conceptualization). However, this same study expresses concern over possible item response set and, if present, "would give spuriously high internal consistency indices" (Sims et al., 1986, p.759).

No stability information is provided by the LSI II. However, Sims et al. (1986) report test-retest reliabilities for the four learning modes ranging from .24 (Concrete Experience) to .66 (Reflective Observation) at fifteen weeks. Sims et al. (1986) also report an increase in the number of subjects who changed learning style classification with retesting when compared with the LSI I and conclude: "Despite its improvements in internal consistency, the LSI II remains unstable across time" (p.756).

Validity

Validity evidence offered by the LSI II (1985) takes the form of a graph depicting the four learning style quadrants with career fields (e.g., journalism, history, data processing) plotted in them. Also provided are learning mode and difference score correlations between the LSI II and the original LSI I (ranging from .87 to .93), and intercorrelations among the learning mode and difference scores of the LSI II. These intercorrelations generally are in the expected direction (negative) and vary widely in magnitude, ranging from -.05 to -.85 with an absolute value mean of .36.

Interpretation

It would appear that the LSI II's gain in internal consistency, over the LSI I, has been nullified by the loss of stability. This, coupled with the general lack of validity information, makes the interpretation of LSI II scores problematic. Until further study of this instrument's psychometric qualities is undertaken, use of this instrument's scores for individual interpretation or advisement, as well as group description, should be considered with even greater care than LSI I scores (described above).

MYERS-BRIGGS TYPE INDICATOR (1985)

The Myers-Briggs Type Indicator (MBTI) is another broadly utilized instrument. The MBTI was designed "to make the theory of psychological types described by C. G. Jung

(1921/1971) understandable and useful in people's lives" (MBTI:Manual, 1985, p.1). The MBTI purports to measure

> "...four basic preferences which, under Jung's theory, direct the use of perception and judgment. The preferences affect not only what people attend to in any given situation, but also how they draw conclusions about what the perceive." (p.2)

The four basic preferences result from an individual's score on four dichotomous indices: Extraversion-Introversion, Sensing-Intuitive, Thinking-Feeling, and Judgment-Perception. Using the combinations of the four preferences sixteen "types" are possible, each including a preference from each indices.

The Extraversion-Introversion (EI) index reflects an individual's preference for orienting her/himself toward the outer or inner world, respectively. Extroverts tend to orient themselves to the outside world of people and objects. Introverts, in contrast, tend to focus on the inner world of concepts and ideas. The Sensing-Intuition (SN) index addresses itself to an individual's preference in the way they perceive. A Sensing person prefers to rely primarily on the observable facts, while an Intuitive individual prefers the realm of relationships or meanings drawn intuitively. The third index, Thinking-Feeling (TF), reflects a person's preference for judging. A Thinking person prefers to utilize logical consequences in making an impersonal decision. In contrast, a ~~Sensing~~ Feeling preference reflects decision making based upon personal or social values. The Judgment-Perception (JP) index reflects a person's preference for dealing with the outside world, with "J's having a more planned, orderly way of life and P's being more spontaneous" (DeVito, 1985).

Instrument Development/Format

The goal of the MBTI is to sort people into groups which Jungian type theory suggests that they already belong to. This is done by utilizing the above described type preferences (i.e., EI, SN, TF, & JP) which are assumed, theoretically, to be four separate dichotomies. In the construction of each preference type index, concern was given to the discrimination of individuals at the continuum's dividing point in order to increase accuracy of type classification.

Item formats are of forced choice type and are either phrases followed by statement choices or a question followed by single word choices; each choice (phrase or word) reflects a type preference. Item generation was by the MBTI authors who "had thoroughly studied Jung's Psychological Types for descriptions and subtle clues about type preferences and their interactions" (MBTI:Manual, 1985, p.141).

Items were then repeatedly administered and revised using a "criterion group" which consisted of individuals who, according to the authors, clearly demonstrated the type preferences under study. Item analysis consisted of item-type correlations across all preference types. Weightings were also developed to account for items' social desirability and sex differences. Further item analysis, with changes in the weighting of items, has taken place with the introduction of revised forms (e.g., Form G).

The MBTI can be scored in two ways. First, scoring can take the form of dichotomous classifications for each of the four indices. Using this format, an individual's score on each of the dichotomous preference indices is used to classify him/her into one of the sixteen types, each type having four components corresponding to the four indices (e.g., INTJ, ENTJ, ISTP). The other scoring option, continuous scoring, uses linear transformations of an individual's preference score for each of the four indices. By using continuous scores, statistical sophistication can be increased in analysis of scores.

Reliabilities

The MBTI Manual (1985) provides a wealth of statistics concerning internal consistency and stability for preference indices within forms F and G, whole forms F and G, across many samples, and continuous and dichotomous scoring. The samples listed for internal consistency cover a wide range of education levels (seventh grade to college graduates, including high school dropouts), ages (9 to 60+ years), aptitude-form F only-(IQ <= 98 to IQ >= 166), and sex.

Reported internal consistency of the continuous scores, using split-half with Spearman-Brown correction, for the preference indices range from the low to mid .80's for pooled sample categories. The lowest coefficients appear in form F with an underachieving seventh grade female sample (.19) and a sample of female subjects in grades seven through twelve with an IQ <= 98 (.41): no similar break-down is given for form G. Coefficient alpha is also calculated for the preference indices, using continuous scores, fewer samples, and only for form F. These alphas range from .64 to .85.

Internal consistency data for dichotomous scoring of type categories is provided by phi and tetrachoric correlations. Phi coefficients range from .49 to .79. Tetrachoric coefficients ranged from .66 to .93. Both statistics are supplied because of the difficulty in MBTI category classification. Tetrachoric correlations assume a normal score distribution (Nunnally, 1978, p.136) and phi assumes true categorical data (MBTI:Manual 1985). By supplying both coefficients the Manual (1985) states that the phi can be used as the low estimate and tetrachoric as the high estimate, with the "true" correlation falling somewhere between.

Stability coefficients for type categories and continuous scoring are also given over various sample groups and forms F and G. Coefficients for continuous scoring range from .86 to .89 for a one week interval to .45 to .58 for a 4 year interval. The 4 year reliability coefficients are depressed by use of a homogeneous sample (medical students) experiencing an attrition rate of 24.5%. Test-retest data concerned with the agreement of type categorizations is given by percentage of agreement between administrations. Time intervals ranged from 5 weeks, with percentages ranging from high 70's to low 90's, to 6 years, percentages ranging from 83% to 90%. In a study using male and female college students, reliabilities ranged from .56 (Thinking -Feeling) to .87 for a 7 week interval (Carskadon, 1977). Lower test-retest reliabilities than reported in the MBTI were also found by Stricker and Ross (1964) who report test-retest reliabilities ranging from .48 (Thinking-Feeling) to .73 over a 14 month interval.

Validity

According to the MBTI Manual (1985), type classification relates to educational achievement in three ways: aptitude, application and interest. Further, the MBTI states that:

> "Academic achievement requires the capacity to deal intensively with concepts and ideas, which are mainly the province of introversion. It also requires the capacity to work with abstraction, symbols, and theory, which are the province of intuition." (Manual, 1985. p.96)

The prediction is then that Introverts-Intuitive (IN) should perform better than Extroverted-Sensing (ES). Also predicted is that Thinking and Feeling types should do well in areas requiring understanding of human motivational factors. Validity data provided, using a variety of aptitude measures, supports this aptitude prediction with the intuition type preferences generally performing better than sensing type preferences. The data comparing introversion and extrovert type preference also is generally supportive of MBTI predictions, although less decisive.

Other validity data is also, in general, supportive. Sewall (1986) reports studies which found agreement between Jungian analysts who type themselves and their MBTI classifications (Bradway, 1964, as cited in Sewall, 1986), and spouses predictions of their mate's MBTI type scores (Cohen, Cohen, and Cross, 1981, as cited in Sewall, 1986). However, some studies report lack of validity evidence with the Extroversion-Introversion and Judging-Perceiving scales (Stricker and Ross, 1964; Cohen, 1978 and Mendelsohn, 1965, as cited in Sewall, 1986).

Another threat to validity are studies by Carlyn (1977) and Richek (1969) which report consistent, positive, and significant intercorrelation of the Sensing-Intuition and Judging-Perception indices.

Interpretation

Worthy of attention, when interpreting MBTI scores, is the lack of bimodal distributions on scales of preference types which are theoretically dichotomous and would, therefore, logically result in bimodal distributions (Stricker & Ross, 1964). Although this lack of bimodal distributions may be an artifact of scale construction designed for maximal discrimination at the type preference dividing point, it complicates the choice of dichotomous or continuous scoring. As DeVito (1975) points out, the result of such distributions makes the choice of whether to use the dichotomous type classification or the continuous score an important one.

Furthermore, of some concern is the choice of samples utilized for item analysis in form revisions. For example, in the development of form G the population tested consisted of students in "grades four through twelve in three public schools in Bethesda, Maryland, and in four private schools in the suburbs of Philadelphia" whose mean IQ was 117; a full standard deviation above the mean (MBTI:Manual, p.144). Further analysis included rescoring and weighting of a sample of University of Florida freshmen. Although reweighted and rescored item/type correlations were generally acceptable, the restrictiveness of the utilized sample leaves questions of item selection bias. Such item selection bias might explain the low reliabilities with low aptitude subjects.

Lastly, the studies finding a lack of independence between the Sensing-Intuitive and Judgment-Perception scale (Carlyn, 1977; Richek, 1969) do not necessarily invalidate the theoretical foundation of the MBTI, they do, however, raise questions concerning the JP scale's inclusion in the instrument and the validity of its interpretation.

However, given the large amount of research data which finds support for the MBTI's reliability and validity, use of MBTI scores appears appropriate with awareness of conflicting research findings. Caution is advised as a result of the above mentioned research findings, and present authors' concerns, which report theoretically inconsistent internal and stability coefficients and a lack of clarity concerning the MBTI's validity. Also, because of reported low reliabilities with low aptitude subjects, pilot testing of the instrument before use with these samples is warranted.

PRODUCTIVITY ENVIRONMENTAL PREFERENCE SURVEY (1982)

The Productivity Environmental Preference Survey (PEPS) was designed to analyze "the conditions under which an adult is most likely to produce, achieve, create, solve problems, make decisions, or learn" (Price, Dunn, & Dunn, 1982, p.1). By studying an individual's PEPS profile, the user is aided "in prescribing the type of environment, working conditions, activities, and motivating factors that would maximize individual output" (p.1). The theoretical basis for the PEPS is not clearly defined and it would appear, as a result, to be atheoretical in nature. The PEPS manual goes further to suggest that characteristics that the user finds related to performance can become criteria for advancement or screening. However, cautions are given concerning the limitations of the instrument-i.e., it does not measure underlying psychological attributes-but "yields information concerned with the patterns through which the highest levels of productivity tend to occur" (p.2). The PEPS reveals how, not why, an individual prefers to learn.

Instrument Development/Format

Construction of the PEPS originated with the identification of "research variables that appeared to describe the ways individuals prefer to learn or work" (Price, Dunn, & Dunn, 1982, p.18). Information concerning what these "research variables" are or how they were identified is not supplied. Items were then constructed to assess these variables and were administered to a sample, the composition of which is not reported. Information concerning the generation of the original items is also not reported. Responses gathered were then factor analyzed and thirty-one variables were identified. A revised version was then administered to "a non-random sample of 589 adults from several states and from various academic and industrial settings" (p.18): further information regarding this sample is not reported. The thirty-one factors were reduced to the twenty-one areas included in the PEPS by eliminating or combining factors which were found to overlap or form a continuum. The 21 elements of the PEPS are:

1. Sound
2. Light
3. Warmth
4. Formal Design
5. Motivated/Unmotivated
6. Persistent
7. Responsible
8. Structure
9. Learning Alone
10. Peer-Oriented Learner
11. Authority-Oriented Learner
12. Several Ways
13. Auditory Preferences
14. Visual Preferences
15. Tactile Preferences
16. Kinesthetic Preferences
17. Requires Intake
18. Evening/Morning
19. Late Morning
20. Afternoon
21. Needs Mobility

There are 100 items which are answered on a Lickert scale of 1 (strongly disagree) to 5 (strongly agree). Raw scores for each of the 21 scales are converted to standardized scores having a range from 20 to 80, a mean of 50, and a standard deviation of 10. A standard score of 40 or less indicates that the individual does not prefer the element "when they work or study" (p.5), while a standard score of 60 or higher indicates a preference for the element. Therefore, only scores which are at least one standard deviation from the mean are used for interpretation. Group summary profiles and area summaries are also available for identifying groups of individuals having similar preferences.

The Manual (1982) reports that the PEPS was originally a true/false instrument and was later changed to its current Lickert format (p.21). It is unclear whether the presented reliability and validity data is from the revised Lickert format or the earlier true/false format. Also, the Persistent items were "revised in an attempt to achieve a more even distribution and to improve the discriminating power of the items" (p.20). However, it appears that the the reported reliability coefficients predate the revision.

Reliabilities

Reported reliability coefficients (Price, Dunn, and Dunn, 1982) are generally low and uneven across the 21 PEPS elements. Reliabilities (Hoyt) range from a low of .29 (Persistent) to a high of .87 (Auditory Preferences), with a mean reliability of .66. Authors report a strong response set, due to social desirability, of certain items. For example, the authors report that "95 percent indicate they complete tasks related to their work" (Price, Dunn, & Dunn, 1982, p.20). Such homogeneity apparently contributed to the Persistent element having the lowest PEPS element reliability. The Motivation/Unmotivated and Kinesthetic Preferences elements appear to suffer from response set attenuation of variance and resulting low reliability. No measures of PEPS reliability over time are reported by the manual.

Validity

Validity provided by the manual is minimal. Eighty-nine of the intercorrelations among the 21 elements are significant at the .05 level, suggesting that the elements are not independent.

Research studies provided by the manual are descriptive in nature and often do not add to PEPS validity. For example, in a study using a college student sample (N = 27) and examining PEPS scores' relationship to the construct of field dependence-independence, no difference was found with regard to preference of structure (Dunn, no date, as cited in Price, Dunn, and Dunn, 1982). However, in another study

by Dunn (no date, as cited in Price, Dunn, and Dunn, 1982) field dependent subjects were found have a stronger preference for structure than field independent subjects.

Interpretation

The manual's lack of clarity concerning the reported instrument data and version makes any interpretation of it usefulness difficult. Therefore, the following discussions concerning PEPS development, reliabilities, and validity are "clouded" by this lack of certainty concerning which PEPS, the true/false or Lickert, the data pertains to.

The consistent lack of detailed information concerning the samples utilized, especially when relatively small, in instrument development is of the great concern. Also, the transformation of raw scores to a standardized score and the decision to use only scores one standard deviation above or below the mean are not explained. The complete lack of stability measures and low internal consistency leaves use of the PEPS, from the position of reliability, extremely questionable and in need of research addressing this issue.

Recommendations for each scale's polarities are given in the form of suggestions for physically or "psychologically" altering the individuals work and/or learning environment. These recommendations are given without references to their source, giving the naive reader the impression of "truth". Without research which verifies these recommendations they must be considered as having been derived from common sense. Such unverified statements do not belong in a learning style instrument manual.

Research which utilizes the PEPS is provided by the manual. Overall, the studies are exploratory in nature and inconclusive. Given the above mentioned reliability concerns and the threat to element independence, little evidence of PEPS validity is present at this time.

Without further knowledge of reliability, validity, and instrument development, interpretation of scores, for groups as well as for individuals, is not warranted without pilot testing and analysis of the instrument on a similar population sample to that of the proposed study.

SELF-DIRECTED LEARNING READINESS SCALE (1977)

No manual is provided for the Self-Directed Learning Readiness Scale (SDLRS). However, the theoretical basis for the SDLRS might be assumed to be founded on the work of authors such as Knowles, Knox, Tough, Houle, and Kidd. Typical assumptions of adult education literature about the adult learner's characteristic include:

-self-directedness
-possession of a wealth of pertinent experience
-desiring immediacy of application
-wanting problem-centered instruction
-wanting a collaborative, informal learning climate

-wanting a role in establishing
-course objectives,
-course planning

The adequacy of these assumptions is debatable (Danis & Tremblay, 1987; Tracy & Schuttenberg, 1986; Check, 1984; Rosenblum & Darkenwald, 1983).

Instrument Development/Format

The Self-Directed Learning Readiness Scale (SDLRS) was developed as part of the doctoral dissertation by Lucy Guglielmino (1977). A modified Delphi technique, involving 14 authorities on self-direct learning, was used to identify and rate important characteristics of self-directed learning. The SDLRS is a self-report questionnaire utilizing a Lickert format. Eight factors of self-direct learning are assessed by the SDLRS: 1) openness to learning opportunities, 2) self-concept as an effective learner, 3) initiative and independence in learning, 4) informed acceptance of responsibility for one's own learning, 5) love of learning, 6) creativity, 7) future orientation, and 8) ability to use basic study skills and problem solving skills. The SDLRS consists of 58 items and is available on floppy disk for Apple II, II+, IIe, and IIc computers. All scoring is done by returning the SDLRS questionnaires to "Guglielmino and Associates".

Reliabilities

A reliability coefficient of .87 is reported for the SDLRS (Guglielmino 1977, as cited in Torrance & Mourad, 1978, and Okabayashi & Torrance, 1984). No further data concerning SDLRS reliability was found in the review of literature or provided by the instrument developers.

Validity

A review of the literature revealed that hypothesized relationships were often found to be nonsignificant (Long & Agyekum, 1983). In contrast to this finding, Firestone (1984) did find a significant correlation between observer ratings of subjects' self-directedness and SDLRS scores.

There also appears to be some evidence of validity for the eight components of the SDLRS, although conflicting. In a study subjecting the SDLRS to principal component analysis, it was found that an eight component solution accounted for 39.4% of the total variance, but that the eight factors derived in this study are not exactly the same as proposed by Guglielmino (Mourad & Torrance, 1979). Mourad & Torrance (1979) qualify their support for SDLRS's validity by stating that the study's sample consisted of 684 gifted students with a mean IQ of 133.06 and that such analysis can be expected to vary with sample variance. Also, some authors have correlated the SDLRS with other

measures which, according to these authors, also measure facets of self-directed learning (Torrance & Mourad, 1978; Mourad & Torrance, 1979; Long & Agyekum, 1983).

Interpretation

Given the lack of data concerning the reliability of the SDLRS, it was not surprising to find little, weak, and conflicting validity data.

A disturbing observation in the literature was that even when authors found little or conflicting validity support for the SDLRS they continued to express support for the SDLRS's validity (Mourad & Torrance, 1979; Long & Agyekum, 1983 & 1984). Authors citing validity evidence on the basis of correlations with other measures of self-directed learning often used measures whose reliability and validity information is also of unkown or questionable quality and their theoretical relationship to the SDLRS is often not fully explained.

The general lack of information concerning the SDLRS's psychometric qualities and the conflicting results of the research make the use of the SDLRS, like the PEPS, not warranted without pilot testing and analysis of the instrument on a similar population sample to that of a proposed study.

PERSONAL STYLE INDICATOR (1985)

Cotroneo's Personal Style Indicator (PSI) Manual (1985) provides the theoretical rationale as, "consistent with Roger's belief in the need for adaptive learning skills" (Cotroneo, 1985, p.12). This rationale results in Cotroneo (1984, as cited in Cotroneo, 1985) suggesting five personal style dimensions for the PSI: Affiliative, measures a person's ability to deal and work with people; Creative, deals with the generation of ideas and comfort in unstructured situations; Enterprising, taps the person's willingness to take risks; Ordered, deals with an individual's ability to organize; and Investigative, a measure of analytical style and the ability to think rationally.

Modeled after the "Strong Vocational Interest Blank" and the "Holland Self-Directed Search" the Personal Style Indicator (PSI) "was designed not as a diagnostic or prescriptive tool, but as an interactive device which relies on the power of interpersonal involvement. Hence, the purpose of the inventory is not to diagnose or prescribe, but to raise awareness and bring people closer" (Cotroneo, 1985, preface) to provide a "springboard to discussion, interpersonal inquiry and increased human understanding" whose goal is to be used "as an interactive tool" (p.15).

Instrument Development/Format

Scale development was guided by the six dimension model utilized in both the "Strong Interest Inventory Blank" and the "Holland Self-Directed Search". Using this model, forty words were generated. Further detail as to how the items were generated is not reported. These forty words were placed in an eight group rank order format and subjected to factor analytic and correlational techniques. The sample used for this analysis is not defined. As a result of analysis, the six original dimensions were condensed to the five PSI dimensions. The five dimensions, described above, are Affiliative, Creative, Enterprising, Ordered, and Investigative. Further correlational analysis then took place to "refine the five PSI scales until statistics of significant and satisfactory magnitude were obtained" (Cotroneo, 1984, p.34). The sample used for this final analysis was made up of college undergraduates, graduate students, and instructors from three southern universities.

The result is an instrument short in length (8 items), simple in format (forced choice rankings of 1 to 5 for each item), and can be hand scored.

Reliabilities

Internal consistency data is presented in the form of five item scale intercorrelation matrices. Overall, these "item scale correlations ranged from .31 to .69 with 87.5% of the correlations being above .50" (Cotroneo, 1985, p.38). The author offers these correlations as evidence of scale homogeneity.

Stability over one and one half months is reported as .83 for the PSI, with the five item scale correlations ranging form .73 to .95.

Validity

Validity data presented by Cotroneo (1985) is consistent with the instrument's predictions, although not extensive or complete. For example, Cotroneo reports that education majors have significantly higher Affiliative scores than do engineering, management science, and accounting/finance majors: analysis consisted of one-way ANOVAs and an undisclosed multiple comparison procedure. This is offered as consistent with findings often found with other measures concerning the interpersonal nature of education majors. Cotroneo continues this procedure of One Way ANOVAs and sample comparison for each scale dimension. In general, the results of these findings are supportive of the PSI's validity, although some findings' interpretations are ambiguous; e.g., science majors are more Ordered than engineering and management majors.

Further convergent and divergent validity information is provided in the form of significant correlations between

satisfaction indices and PSI dimensions seen as important to occupation (Cotroneo, no date, as cited in Cotroneo, 1984). These are presented to demonstrate the PSI's ability to differentiate among individuals in a theoretically predictable manor.

Interpretation

In general the reliability of the PSI is of acceptable strength. However, the lack of literature which addresses the reliability on diverse samples gives cause for cautious interpretation. For example, use of a homogeneous sample of thirty engineering students may have actually attenuated the reported stability correlations.

The ability of the PSI to differentiate between occupations on the basis of style differences and its correlation to job satisfaction adds to the PSI's claim to validity. However, caution should be used when interpreting the correlational results of the study because the author does not give any information concerning the job satisfaction instrument's reliability or validity.

Overall, the PSI appears to have a well conceived theoretical base and to be well constructed. However, lack of detailed information concerning the instrument's construction and the lack of normative data, pertaining to reliability and validity, requires the user to exercise caution when trying to interpret group or individual scores. The author appears cognizant of these limitations and cautions the user repeatedly in the manual. In order for the PSI to gain interpretive strength, further research is necessary and would appear (a priori) to be promising.

GROUP EMBEDDED FIGURES TEST (1971)

The Group Embedded Figures Test (Oltman, Raskin, & Witkin, 1971, as cited in Witkin, Oltman, Raskin, & Karp, 1971) is an instrument designed to measure the construct of field-dependence/field-independence. The construct of field-dependence/field-independence (FD/FI) is used to describe the systematic differences in perceptual functioning. This measure of disembedding is theoretically linked to cognitive differentiation. Witkin et al. (1972) found that individuals differed in their spatial orientation and that this orientation is linked to an extensive and varied array of other personality characteristics. Among the array of characteristics found to be related to FD/FI construct are an individual's problem-solving activities, memory recall, and vocational choice/performance (Witkin, et al., 1962, 1977).

Instrument Development/Format

The Group Embedded Figures Test (GEFT) is a group administered version of the individually administered

Embedded Figures Test (EFT). The GEFT's popularity, as a FD/FI measure, lies in its short administration time (20 minutes), applicability for testing large groups of subjects at once, and ease of hand scoring. The GEFT consists of 25 items, of which only 18 are actually scored. The GEFT is a timed test in which the subject is required to locate, in each of 25 items, a simple geometric figure embedded in a more complex figure. The GEFT is broken down into three separately timed sections: 1) lasts 2 minutes and contains 7 unscored items, 2) & 3) each lasts 5 minutes and contains 9 scored items. The ability to locate the embedded figures is characteristic of field-independence.

Reliabilities

Because the GEFT is a speed test, internal consistency was measured by treating each scored section (2 & 3) as split-halves. The Spearman-Brown correction produced a reliability coefficient of .82. Carter and Loo (1980), using a sample of 266 undergraduates, report an internal consistency coefficient (Cronbach's alpha) of .86 for both males and females on the whole instrument. Panek, Funk, and Nelson (1980) report somewhat low results using split-half techniques with the reliabilities ranging from .57 (ages 25 to 32) to .90 (ages 33 to 40); mean of .75 covering ages 17 to 72.

No stability coefficients are reported in the manual. The test-retest reliability of the EFT is reported as .89 over a three year interval (Bauman, 1951, as cited in Witkin, Oltman, Raskin, & Karp, 1971).

Validity

Criterion-related measures of validity lie in the relationship between the GEFT and its "parents": Embedded Figures Test and the Rod and Frame test. Reported correlations between the GEFT and the EFT are -.82 for men (73 male undergraduates) and -.63 for females (68 female undergraduates). Negative correlations are the result of differences in the scoring procedures between the two measures. Also reported are the correlations between the GEFT and the Portable Rod and Frame Test (PRFT): -.39 for males and -.34 for females. Since the Rod and Frame Test "is itself a criterion measure of field-dependence-independence" (Witkin, Oltman, Raskin, & Karp, 1971, p.28), these low correlations are not supportive of the GEFT validity. A low correlation, -.46, is also reported by Panek et al. (1980).

Interpretation

Use of GEFT is cautioned by the authors.

> "The combined evidence suggests that the GEFT may prove to be a useful substitute for the EFT when individual testing is impractical. It must still be considered a research instrument, however, until more extensive direct and construct validation data are collected." (Witkin, Oltman, Raskin, & Karp, 1971,p. 29)

Given the high internal consistency and stability correlations, although only one stability coefficient was found in the literature, the GEFT appears to be a reliable instrument of questionable validity. The authors' caution concerning the use of the instrument is therefore warranted. Because of it ease of administration and high reliability, research aimed at improving its validity seems appropriate. Until such time, interpretation of scores at the individual level appear inappropriate and group scores should only be used as a gross measure of the field dependence-independence construct.

CONCLUSIONS

The reviewed instruments come from varied theoretical perspectives, from Jungian type theory (Myers-Briggs Type Indicator) to Rogerian theory on adaptive learning (Cotroneo's Personal Style Indicator), and range from single element tests (Witkin et al. Group Embedded Figures Test) to the extremely multi-faceted (Price et al.'s Productivity Environment Preference Survey). Also, instrument development methodologies range from simplistic to sophisticated, and reliability and validity evidence ranges from barely existent to well developed.

Some instruments (e.g., MBTI & GEFT) have been been in use for a long time and others are relatively new. The above discussions have demonstrated that length of existence does not necessarily mean the instrument is any more valid or reliable than its newer counterparts. What the older instruments do have is the vast amount of available information (i.e., research articles and normative data) that, when properly examined, provides the basis for informed choices concerning the instrument use; e.g., theoretical "fit", reliability and validity for the proposed population sample, and how best to interpret results.

Although the newer instruments (e.g., PEPS, SDLRS, LSI I-II, & PSI) offer plenty of opportunity for validation and reliability studies, insufficient data is available for the user to make informed decisions. For these newer instruments to become the "state-of-the-art", much more <u>quality</u> information is needed.

Unfortunately, many users of learning style instruments are psychometrically naive and may be overwhelmed from the apparent current trend to merchandise test of learning style before reliability and validity is established. The typical consumer-teacher, administrator, manager, and parent too often accept the claims of sales brochures without question.

The most obvious negative outcomes are that inappropriate decisions may be made about individuals, and that the legitimate study of learning style may also be corrupted by the acceptance of instruments with unverified, in not lacking, validity.

SUGGESTIONS FOR FUTURE RESEARCH

Despite the critical critique of learning style instruments present above, the present authors agree with Sewall's (1986) call for further exploration of the relationship between learning styles and actual learning environments or particular instructional techniques.

However, seen as having a greater immediate need is the continued psychometric refinement of the learning style instruments. Until the learning style instruments used for examining adult learner characteristics are refined, the exploration of the construct will continue to be produce conflict and often uninterpretable results. Of concern is the fact that many of the conflicting results reported in the current literature may be artifacts of the learning style instruments used in the study.

As an aid to this call for refinement the question must be asked, "What is the construct of learning style?" One possible procedure, out of many, for examining the construct and "its" instruments is the utilization of multi-method-multi-trait techniques. Such research needs to encompass diverse samples and directly address the problem of convergent and divergent validity. By exploring adult learning style with such techniques and scope, the definition of learning style can become better defined, resulting in greater construct parsimony. With (a) refined definition(s), learning style instruments can be constructed and refined for both researcher and practitioner, thus allowing clearer interpretation of research results and scores used for guiding adult learners in academic and vocational decision making.

REFERENCES

Carlyn, M. (1977). An assessment of the Myers-Briggs Type Indicator. Journal of Personality Assessment, 41(5), 461-473.

Carskadon, T. G. (1977). Test-retest reliabilities of continuous scores on the Myers-Briggs Type Indicator. Psychological Reports, 41, 1011-1012.

Carter, H., & Loo, R. (1980). Group Embedded-Figures Test: Psychometric data. Perceptual and Motor Skills, 50, 32-34.

Check, J., (1984). Teaching learning preferences of the adult learner. Education, 105(1), 107-114.

Cotrone, K. (1985). Understanding personal style: A manual for the Personal Style Indicator. Athens, GA: Academic Excellence Leadership Project.

Danis, C., & Tremblay, N. (1987). Propositions regarding autodiactic learning and their implications for teaching. Lifelong Learning, 10(7), 4-7.

DeVito, A. J. (1985). Review of Myers-Briggs Type Indicator. In J. Mitchell (ed.), The ninth mental measurements yearbook (pp.1030-1032). Lincoln, NE: University of Nebraska Press.

Firestone, P. M. (1984). A construct validation of the Self-Directed Learning Readiness Scale with labour education participants. Dissertaion Abstracts International, 46(5), 1160-1161A.

Fox, R. D. (1984, Winter). Learning styles and instructional preferences in continuing education for health professionals: A validity study of the LSI. Adult Education Quarterly. 35(2), 72-85.

Freedman, R., & Stumpf, S. (1978). What can one learn from the Learning Style Inventory? Academy of Management Journal, 21(2), 275-282.

Geller, L. M. (1979). Reliability of the Learning Style Inventory. Psychological Reports, 44, 555-561.

Holtzclaw, L. R. (1985, April). Adult learners' preferred learning styles, choice of courses, and subject area for prior experiential learning credit. Lifelong Learning, 23-27.

Kolb, P. A. (1976). Learning Style Inventory: Technical Manual. Boston: McBer and Company.

Korhonen, L. J., & McCall, R. J. (1986). The interaction of learning style and learning environment on adult achievement. Lifelong Learning, 10(2), 21-23.

Learning-Style Inventory: Technical Specifications. (1985). Boston: McBer and Company.

Long, H. B., & Agyekum, S. K. (1983). Guglielmino's Self-Directed Learning Readiness Scale: A validation study. Higher Education, 12, 77-87.

Long, H. B., & Agyekum, S. K. (1984). Teacher ratings in the validation of Guglielmino's Self-Directed Learning Readiness Scale. Higher Education, 13, 709-715.

Marshall, J. C., & Merritt, S. L. (1985). Reliability and construct validity of alternate forms of the Learning Style Inventory. Educational and Psychological Measurement, 45, 931-937.

Mourad, S. A., & Torrance, E. P. (1979). Construct validity of the Self-Directed Learning Readiness Scale. Journal for the Education of the Gifted, 3,(2), 93-104.

Myers, I. B., & McCaulley, M. H. (1985). Manual: A guide to the development and use of the Myers-Briggs Type Indicator. Palo Alto: Consulting Psychologists Press.

Nunnally, J. C. (1967). Psychometric theory. New York: McGraw-Hill.

Okabayshi, H., & Torrance, E. P. (1984). Role of style of learning and thinking and self-directed learning readiness in the achievement of gifted students. Journal of Learning Disabilities, 17(2), 104-107.

Panek, P. E., Funk, L. G., & Nelson, P. K. (1980). Reliability and validity of the Group Embedded Figures Test across the life span. Perceptual and Motor Skills, 50, 1171-1174.

Pigg, K. E., Busch, L., & Lacy, W. B. (1980). Learning styles in adult education: A study of county extension agents. Adult Education, 30(4), 233-244.

Price, G. E., Dunn, R., & Dunn, K. (1982). Manual: Productivity Environmental Preference Survey. Lawrence, KS: Price Systems.

Richek, H. G. (1969). Note on intercorrelations of scales of the Myers-Briggs Type Indicator. Psychological Reports, 25, 28-30.

Rosenblum, S., & Dorkenwald, G. G. (1983). Effects of adult learner participation in course planning on achievement and satisfaction. Adult Education Quarterly, 33(3), 147-153.

Sewall, T. J. (1986). The measurement of learning style: A critique of four assessment tools. Green Bay: Wisconsin Univ., Assessment Center. (ERIC Document No. ED 267 247)

Sims, R. R., Veres III, J. G., Watson, P., & Buckner, K. E. (1986). The reliability and classification stability of the Learning Style Inventory. Educational and Psychological Measurement, 46, 753-760.

Striker, L. J., & Ross, J. (1964). An assessment of some structural properties of the Jungian personality typology. Journal of Abnormal and Social Psychology, 68(1), 62-71.

Sugarman, L. (1985). Kolb's model of experiential learning: Touchstone for trainers, students, counselors, and clients. Journal of Counseling and Development, 64, 264-268.

Torrance, E. P., & Mourad, S. (1978). Some creativity and style of learning and thinking correlates of Guglielmino's Self-Directed Learning Readiness Scale. Psychological Reports, 43, 1167-1171.

Tracy, S. J., & Schuttenberg, E. M. (1986). Exploring adult learners' rationales for course interaction preferences. Adult Education Quarterly, 36(3), 142-156.

West, R. F. (1982). A construct validity study of Kolb's learning style types in medical education. Journal of Medical Education, 57, 794-796.

Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. R., & Karp, S. A. (1962). Psychological Differentiation. New York: Wiley.

Witkin, H. A., Oltman, P. K., Raskin, E., & Karp (1971). Manual: Embedded Figures Test, Children's Embedded Figures Test, and Group Embedded Figures Test. Palo Alto: Consulting Psychologist Press.

Wunderlich, M. D., & Gjerde, C. L. (1978). Another look at Learning Style Inventory and medical career choice. Journal of Medical Education, 53, 45-54.